

# Sentiment Polarity Classification of Corporate Review Data with a Bidirectional Long-Short Term Memory (biLSTM) Neural Network Architecture

R. E. Loke<sup>a</sup> and O. Kachaniuk

*Centre for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands*

**Keywords:** Natural Language Processing (NLP), Sentiment Analysis, Corporate Review Data, Supervised Learning, biLSTM Neural Network with Attention Mechanism, Word Embeddings.

**Abstract:** A considerable amount of literature has been published on Corporate Reputation, Branding and Brand Image. These studies are extensive and focus particularly on questionnaires and statistical analysis. Although extensive research has been carried out, no single study was found which attempted to predict corporate reputation performance based on data collected from media sources. To perform this task, a biLSTM Neural Network extended with attention mechanism was utilized. The advantages of this architecture are that it obtains excellent performance for NLP tasks. The state-of-the-art designed model achieves highly competitive results, F1 scores around 72%, accuracy of 92% and loss around 20%.

## 1 INTRODUCTION

The last two decades have seen a growing trend toward usage of social media. This trend highlighted the need to process huge numbers of client's opinions, reviews, regarding their experiences into meaningful insights which could be used to boost corporate reputation.


In the history of development of economic studies, corporate reputation has been thought of as a key factor in corporate performance. Numerous researches suggest a positive correlation between corporate reputation and, for instance financial performance (Silvija et al., 2017), Gatzert (2015). Furthermore, Keh et al. (2009) show a relation between reputation, customer's purchase intention and the willingness to pay a price premium, where they underline significant influence of customer trust/commitment on those two variables. Ross et al. (1992) argue that socially engaged companies benefit from it by having higher sales.

In the broad use of the term 'Corporate Reputation', it is sometimes equated as an entity consisting of the views and beliefs about the company. It encompasses the past and possible future of corporation. However, the definition of corporate

reputation is a rather nebulous term due to the variety, high polarity, subjectivity and various perspectives of human judgment (Lester, 2009). According to the definition provided by Fombrun and van Riel (1997), corporate reputations are 'ubiquitous, they remain relatively understudied'.

One of the greatest challenges in the task of predicting corporate reputation from media is related to opinion mining, a machine learning task, which refers to topic extraction from all data flow. A precise method that is suited to this task is Attention Based Sentiment Analysis (ABSA) that distinguishes sentiments in a sentence according to aspects, so one sentence can contain sentiment for different aspects (Jiang, Chen, Xu, Ao and Yang, 2019). ABSA methods can be set to work in both supervised as well as unsupervised learning paradigms. In both paradigms, BiLSTM networks with an attention mechanism are increasingly important for this sophisticated machine learning task and applied linguistics as these require advanced text analysis techniques.

In the following sections, we describe method, data, results as well as discussion and conclusion.

<sup>a</sup> <https://orcid.org/0000-0002-7168-090X>

## 2 METHOD

### 2.1 Bidirectional LSTM

Artificial Neural Networks (ANNs) are machine learning algorithms based on the principles and architectures of the human brain. NNs are a successful attempt to mimic functioning and learning ability of humans, where the neurons play a role of computational nodes receiving signals via weighted connections, synapses. Accordingly, the result computed by an activation function and the output can be found on the next node (Lurz, 2018). Figure 1 illustrates an input layer with  $x_1 - x_4$ : input neurons sending information to the next hidden layer. Every neuron has weighted inputs, mentioned in previous paragraph, the synapses. On the next step, weights are summarized and with the help of an activation function pushed to output. There is a variety of activation functions, such as linear, step, sigmoid, tanh, and rectified linear unit (ReLU). In the model that we created ReLU activation function closely connected with a sigmoid output layer was applied.

ANNs can be arranged and connected in multiple different ways according to the goals, constraints and type of data. One of the most effective architectures for solving NLP related tasks is Recurrent Neural Networks (RNNs). It gained wide application for handling sequenced data and time series.

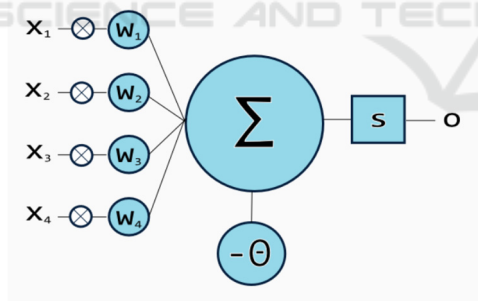


Figure 1: Visualization of Artificial Neuron (Davydova, 2017).

Deep learning is the modern state-of-the-art for certain domains. It has performed to a high standard for natural language processing tasks in the last decade. In support of it, RNN has shown outstanding results in capturing contextual similarities. However, there are certain drawbacks specifically related to this type of ANN architecture. What reduces precision of this type of network is that the order of words matters, in the sense that later words have more weights in analysis (Lai, Xu, Liu and Zhao, 2015). The movement from left to right defines computation of

sequence's probability in one direction. Accordingly, the last words of a sentence have more influence. Therefore, accuracy of prediction at the next time step is lower. However, this limitation can be tackled by using an extended version of traditional RNN: Bidirectional LSTM.

One of the most successful network architectures in state-of-art ANN algorithms, is a subclass of RNN: the bidirectional LSTM model. LSTM differs from RNN in a number of important ways: it does not use an activation function within recurrent components, stored values stay fixed and the gradient vanishing/exploding problem is not relevant. Figure 2 illustrates the basic LSTM model. This shows an extension with the variable,  $c_t$ . The purpose of this variable is to collect the memory about previous time steps and pass it through the network. LSTM units are implemented in blocks having three gates: input, forget and output. This structure provides an additional input to the unit as input from a previous time step,  $h_{t-1}$ , and new input,  $x_t$  (Manning and Socher, 2017).

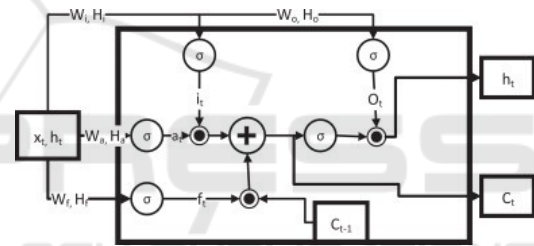


Figure 2: LSTM cell (Kachaniuk, 2019).

LSTM cells pass information from the past outputs to current outputs with the help of storage elements. LSTM has three control signals,  $i$ ,  $o$  and  $f$  respectively. Each of the non-linear functions  $\sigma$  is activated by a weighted sum of the current input observation  $x_t$  and previous hidden state  $h_{t-1}$ . In gate  $f_t$  a decision is made about remembering the previous state, input gate  $i_t$  is responsible to make the decision whether to update the state of the LSTM using the current input or not; output gate  $o_t$  decides to pass the hidden state further to the next iteration (Elsheikh, Yacout and Ouali, 2018).

The new state  $c_t$  stored in the LSTM is the sum of the new gated input at and the gated previous state as shown in Figure 2. This illustrates how the gates and inputs interact.

LSTM proves to be a powerful model achieving high prediction accuracy.  $W$ ,  $H$ , and  $b$  are the trainable weights and biases, for each gating signal indicated, while  $h_t$  is the current hidden layer activation. Similar to the hidden states, the cell states are defined. The current input is  $x_t$ , and the gate

activations are  $f_t$ ,  $a_t$ ,  $i_t$  as described previously. Finally,  $\circ$  is the element-wise multiplication operator (Elsheikh, Yacout and Ouali, 2018).

## 2.2 Word Embedding

Word embedding is fast becoming a key instrument in NLP. The input to the machine learning algorithm consists of letters, words, sentences or documents, accordingly data needs to be mapped in a way that it can be further numerically processed by a machine (Lurz, 2018). The state-of-art word representation that we use is performed accordingly on a word level by using character level information.

An embedding plays a crucial role in mapping of a discrete categorical variable to a vector of continuous numbers for a neural network. It is essential for representation of categorical variables as a meaningful input to a machine learning model. Word embeddings can significantly reduce dimensionality and loss for such models. For instance, the pair of synonyms in the vocabulary that is associated with positive reviews will be placed near each other in the embedding space. The network was trained so that both of those words refer to a positive review. Neural network embeddings are practically useful to learn how to represent discrete data as low dimensional continuous vectors and to depreciate limitations of traditional encoding (Koehrsen, 2018).

It is useful to compare widely used embeddings and give a reasoning for the particular choice of fastText.

In linguistics, a word is defined as the smallest element, which can be expressed 'in isolation with objective or practical meaning' (Wikipedia, n.d.). Worth mentioning, that by meaning is understood what 'a word, action, or concept is all about — its purpose, significance, or definition' (Vocabulary.com, n.d.). Those definitions pose a problem of an accurate representation of a word's meaning for Natural Language Processing (NLP) because of the potential for ambiguity and dependence on context they introduce. This shows a need to be explicit about exactly what is meant by the meaning of a word. Furthermore, meaning of word is an increasingly important area in applied linguistics. Accordingly, British linguist, J.R. Firth (1957) suggests the following formulation: "You should know a word by the company it keeps". This underlines the importance to apply an approach for its representation which encodes the meaning of words in such a way that similarity between words will be easily seen.

Previously, techniques attempted to represent words by using taxonomic resources. Among others, - the main python nltk library - wordnet. This package contains a considerable number of synonyms and relation sets. However, the nature of a word's meaning remains subjective, it is challenging to distinguish word's similarity, and unrealistic to keep a package like wordnet updated. Furthermore, the greater part of rule-based approaches, treat words as an atomic symbol representation. One of the challenges of this approach is that it will not match words with similar meaning like, for instance, "notebook" and "laptop". All mentioned above support an idea about localist representation, "one-hot" vector encoding being an inefficient method which has no inherent notion of relationships between words. Accordingly, loosely described this representation considers a word independent of any context. Those negative features can be implied not only for symbolic encodings but for many other probabilistic, statistical conventional ML approaches.

Overall, these cases support the importance of exploring an approach which can display the meaning of words in such a way that you can directly see similarity between them (Manning and Socher, 2017). This formulation is a commonly-used notion in philosophy, the theory of meaning; more precisely to semantic and foundational theories (Speaks, 2018). Theory of meanings in this context refers to a semantic theory, to a "specification of the meanings of the words and sentences of some symbol systems" (Stanford Encyclopedia of Philosophy, 2014). While a variety of definitions and interpretations have been suggested, a distributional approach will follow the theory of meaning. Levy et al. (2015) claim better performance of predictive models for neural-network-inspired word embedding models than count-based ones. In his research, both GloVe and word2vec embeddings considered to be prediction based, accordingly, outperform traditional models in similarity and relatedness tasks. The findings indicate that hyper-parameters tuning and sufficient amount of data could improve performance of embeddings in general (Koehrsen, 2018). However, GloVe differs from word2vec in a number of respects. GloVe tends to perform better than the word2vec skip-gram model on analogy tasks in combination with matrix factorization methods which can make full use of global statistical information (Pennington, Socher and Manning, 2014).

The main challenge that we face is related to the fact that the language of reviews is Dutch. Recent researches in NLP have heightened the need for use of specific language embeddings in order to increase

efficiency of word form and context representation for improved encoding in word embedding (Qian, Qiu and Huang, 2016). FastText, with specific Dutch language embedding, was introduced for performing aspect-based sentiment analysis. Whereas GloVe/word2vec function on a word level, fastText operates on a character level. Moreover, it processes a word as a composition of these characters n-grams. Thus, a word vector summarizes all the characters n-grams. In contrast to GloVe and word2vec, it can construct the vector for a word even if this word has not been seen previously in the training set data (“out of vocabulary”). And consequently, it generates better embeddings for rare words. In practice, a model used in the experiments for this work had better convergence when fastText was used compared to GloVe (Rajasekharan, 2017). Therefore, it was decided to proceed with fastText embeddings.

Accordingly, pre-trained vectors for Dutch language were added to the biLSTM model as one of the layers. Moreover, these vectors were trained on Common Crawl and Wikipedia using fastText. The model was trained using CBOW with position-weights, in dimension 300, with character ngrams of length 5, a window of size 5 and 10 negatives (Facebook Inc., 2020). This embedding is developed by the Facebook AI team as morphologically enriched word vectors with subword information (Bojanowski, Grave, Joulin and Mikolov, 2017). This approach was initially introduced by Schütze in 1993. The most popular embeddings which can be found back in numerous papers and researches, are the word2vec and GloVe models. Despite common usage of those embeddings, fastText outperforms them in a range of tasks. The accuracy of the GloVe model outperforms models of similar size and dimensionality such as SVD, word2vec (skip-gram and CBOW) for a word analogy task given the percentage for semantic and syntactic accuracy’s (Pennington, Socher and Manning, 2014). Both Word2vec and GloVe generate vectors on a word level, while fastText processes every word as a collection of n-grams. Moreover, fastText can be loosely described as an extension of the skipgram model (Mikolov et al., 2013) in which words are embodied as the sum of n-gram vectors (Bojanowski, Grave, Joulin and Mikolov, 2017). This technique enables preserving morphological meaning of words.

According to limitations of word2vec, it is worth mentioning that this embedding performs learning exceptionally for complete words on training data; while fastText learns its vectors on the n-grams and complete words. To achieve better performance for each vector, improvements are incorporated

uniformly. Therefore, those techniques are relatively time consuming, at every step the average of n-gram should be computed. However, there are several important aspects where fastText makes an original contribution. For instance, word embeddings with character level information tend to be more accurate as they generate better embeddings for rare and oov words (out of vocabulary) (Cesconi, 2017).

In some research (Perone et al., 2018) performance of a number of widely used word embeddings methods was evaluated on a broad range of downstream and linguistic feature probing tasks, among which: multi-class classification, entailment and semantic relatedness, semantic textual similarity, paraphrase detection. In particular, the results of the conducted analysis cannot be unequivocally interpreted. The universal encoder does not exist yet. However, results presented in Kachaniuk (2019) show that fastText slightly outperforms word2vec and GloVe in most of task categories. The paper by the Facebook AI team describes its findings and addresses improvements made compared to skip-gram, CBOW implementations and other word vectors incorporating subword information on a word similarity task.

## 2.3 Attentional Mechanism

An attention mechanism is an algorithm which can allocate attention to more important words in a sentence by adjusting the weights they assign to various inputs. This mechanism was successfully exploited in machine translation (Bahdanau, Cho, and Bengio, 2014), image classification (Wang, Jiang, Qian, Yang, Li, Zhang, Wang and Tang, 2017), and speech emotion recognition (Chorowski, Bahdanau, Serdyuk, Cho and Bengio, 2015).

As was pointed out in the previous section, the context in which a word is used is more important than the meaning of a word itself. Alternatives for attention mechanisms, such as WordNet or sense2vec, do not take into account connections among words in a sentence. They are aimed at a single word representation. Accordingly, the approach implemented in this paper fully tackles the limitations of alternative methods. The relationships of every word in sentence can be calculated with the help of linear algebra. The vector of meaning will contain respectively those expressed relationships. Moreover, an attention mechanism weighs every word in a sentence attempting to assign the highest weight to the word which is the most crucial regarding the context (Nicholson, n.d.).



This work uses a bidirectional RNN. Accordingly, the meaning of a word relates not only to the word in front of it but also the one which is behind. The model architecture of bidirectional LSTM is designed, accordingly, to enable the eigenvector to learn in two directions. The way it is created, enhances its semantic and contextual capacity compared to unidirectional models. The attention mechanism enables the model to learn a weight for each word assigning heavier weights to key-words (Du and Huang, 2018).

Together these results provide important insights into the designed neural network architecture. Taken together, an attention mechanism is a successful attempt to improve understanding of neural network processing with additional intelligence. Hence, an attention mechanism appoints weights according to the input regarding to the problem to be solved (Galassi, Lippi and Torroni, 2019).

### 2.4 biLSTM Model Architecture

The model designed has the following architecture. The first layer of this model consists of the fastText embedding, previously described, with dimensionality of 300, followed by a biLSTM layer with 128 network units for each gate. This model is extended with the attention mechanism which is built in as an additional layer on top of the biLSTM layer. Next to it, 2 fully connected layers: ReLU was used as the activation function for hidden layers of size 224 and sigmoid as the output layer with size of 8 classes. Experiments with other activation functions did not show any improvements for model performance.

## 3 DATA

The research data in this work was scraped from three main online resources: trustedshop.nl, trustpilot.nl and kiyoh.nl. Furthermore, data for this study was collected using Python with a Scrapy Spider and a browser automation tool - Selenium for dynamic pages.

In total approximately 1 million reviews were collected. Most of these reviews were scraped from Trustpilot, as is shown in Figure 3 a.

From the large set of data, a subset of approximately 3,000 reviews was randomly selected for training and validation purposes. Only reviews in Dutch language were processed for this study; English language reviews were strictly omitted from the training and validation data.

The data set was labeled by two experts who assigned 3-class polarity sentiment on a phrase level respectively to 4 predefined aspects that were found to be important in corporate reputation - assortment, product, service and delivery.

Figure 3 b shows the balance in the respective sentiment polarity. What stands out in the figure is that the highest proportion is given to neutral sentiment. A possible explanation for this could be the fact that neutrality indicates also absence of any negative or positive sentiment. Overall prevailing presence of positive sentiment indicates high satisfaction levels with the products and services of this data set. Consequently, negative reviews, the lowest proportion in all categories, will be more difficult to predict as it is easier to predict the majority class of a sentiment. Accordingly, the overall accuracy of the prediction might not be the best estimator for performance of the model. Therefore, other evaluation techniques were applied as well.

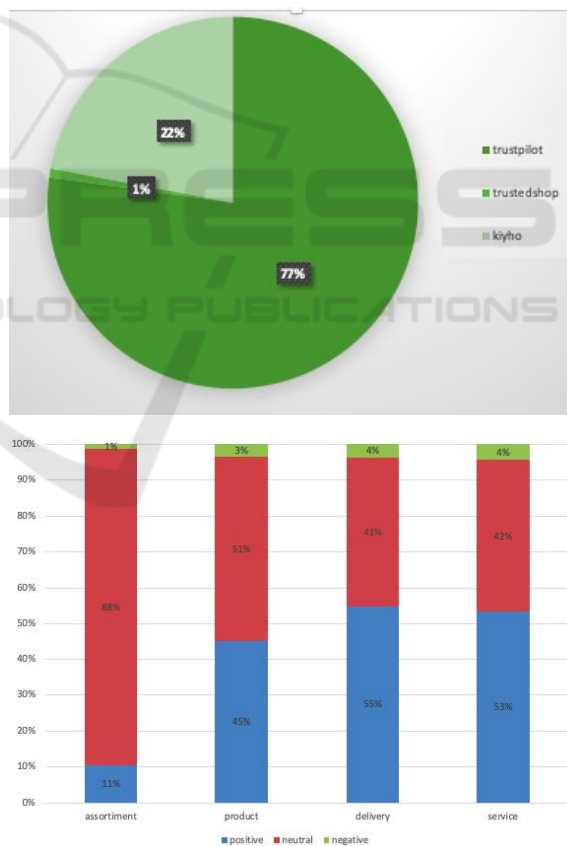
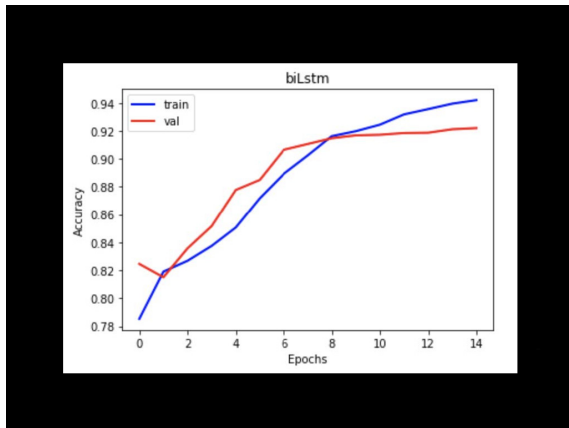


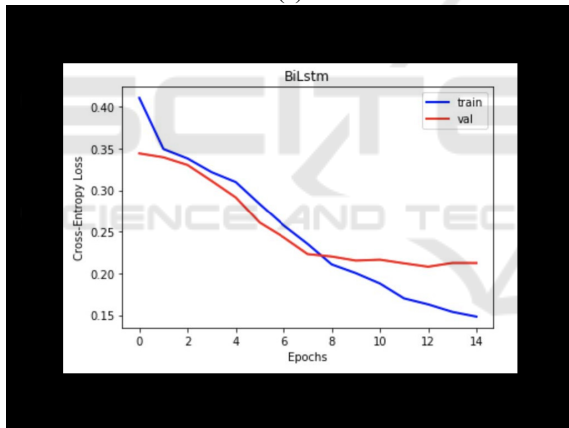
Figure 3: a: Share of number of reviews per review website; b: Breakdown of labeled sentiment in percentages on the aspects assortment, product, service and delivery with blue denoting positive sentiment, red denoting neutral sentiment and green denoting negative sentiment.

## 4 RESULTS

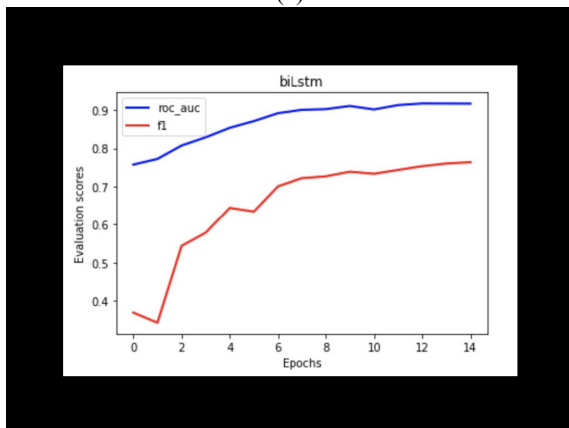
There are a number of instruments available for measuring the performance of neural networks. In most recent studies, it has been measured by looking at such metrics as accuracy, loss function, ROC AUC scores and F1-scores. The final results are shown in Figure 4 a, b and c.



(a)



(b)



(c)

Figure 4: Evaluation scores.

The results obtained for ROC-AUC and F1-scores indicate good performance of the model, however, the question is how reliable those evaluations are taking into account that the problem relates to multi-class classification. In order to understand how predictions for all classes are distributed, an analysis of pair-wise confusion matrices for negative and positive classes for each component was performed. See Figures 5 to 8.

The pairs of matrices are quite revealing in several ways. First, it is obvious that not all classes were as well predicted as it can be seen from previous section. Secondly, the pos-assortment precision is impressively high and recall is low. It indicates that the classifier is specialized in the way that it predicts a small number of labels while data contains plenty of it. In contrast to neg-assortment which is hardly present throughout the labeled reviews. Consequently, precision and recall have 0 values.

Neg-product confirms the conclusion deduced from the previous Figure 5: the number of reviews with negative sentiment for products is low, therefore, it is difficult to predict. In contrast to it, pos-product achieves satisfying performance, recall and precision are high.

The next result, shown in Figure 7, indicates that the classifier predicts both negative and positive service classes with high precision and reasonable recall.

The results obtained for negative delivery are comparable to the class of positive assortment, in the sense that the value of recall is low and precision is around 70%. Figure 8 for negative delivery can be interpreted as follows: most of the negative delivery cases were not predicted, the negative delivery instances were not recognized. The performance of positive delivery is good.

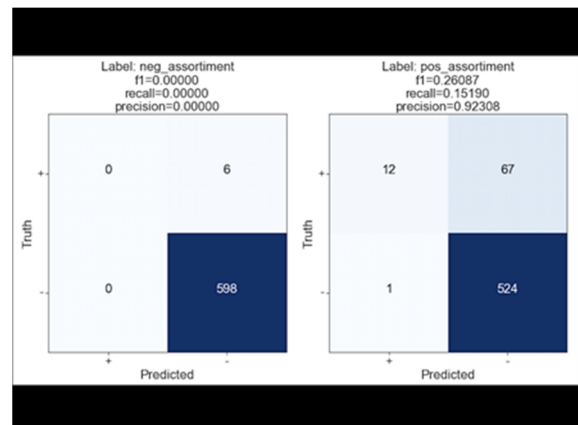


Figure 5: Confusion Matrix for Assortment Class.

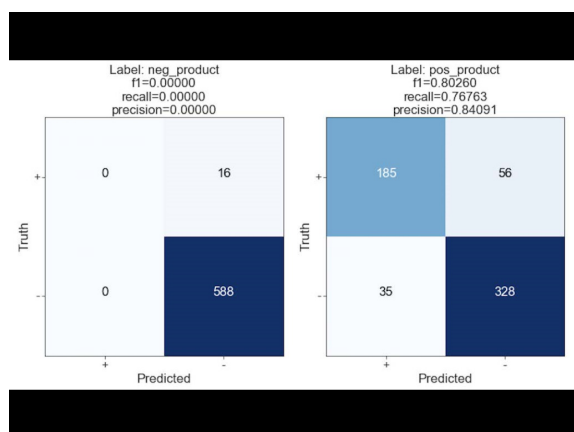


Figure 6: Confusion Matrix for Product Class.

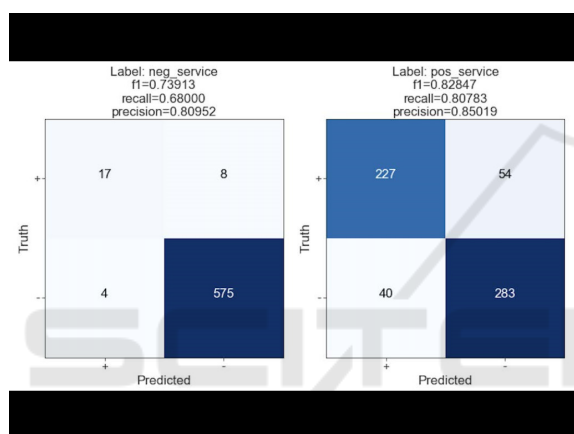


Figure 7: Confusion Matrix for Service Class.

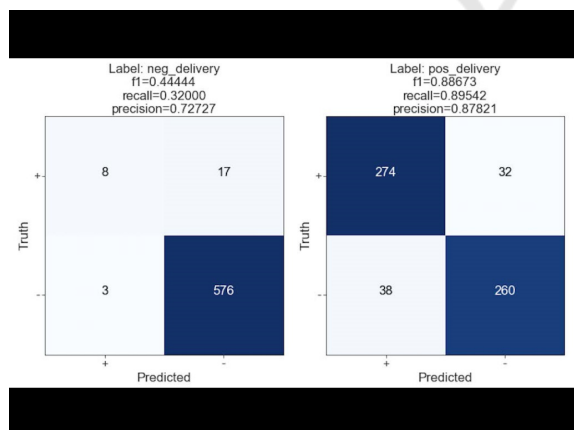


Figure 8: Confusion Matrix for Delivery Class.

## 5 DISCUSSION AND CONCLUSION

In our experiment on the effect of sentiment polarity classification, we found that a biLSTM model equipped with additional attention layer, and word embedding with character information have a clear positive effect and improve overall performance.

The empirical results acquired have shown that a model using biLSTM architecture achieves a high overall F-1 score under the condition that sufficient training data are available. The designed model achieves competitive state-of-the-art results in accuracy of 92%, loss of 20%, F-1 scores of 72% and ROC-AUC of 90%. These results are comparable with other state-of-the-art models.

This indicates that these models have sufficient complexity to learn the morphological and lexical patterns from the annotated online reviews. The most obvious limitation of this work is a lack of annotated data, especially reviews indicating negative sentiment. The small number of negative reviews suggested to us that satisfied clients are more common and/or more frequently tend to share their positive experiences. The data used in this study appear to be unbalanced.

The results of this research support the idea that sentiments related to aspects of corporate reputation can be predicted from online available review websites, applying state-of-the-art neural networks technology. This is important for retail organizations that are aiming to analyze or improve their performance.

## ACKNOWLEDGEMENTS

Oksana Kachaniuk would like to thank Sandjai Bhulai for providing valuable advice and guidance as well as Tibert Verhagen for the fruitful collaboration in relevant projects amongst which Kachaniuk (2019) that made this work possible.

Rob Loke would like to thank the four anonymous reviewers for their valuable comments to the first version of the submitted manuscript.

## REFERENCES

Bahdanau, D., Cho, K., Bengio, Y, 2014. Neural machine translation by jointly learning to align and translate. *arXiv e-prints*, abs/1409.0473, 09. 24

- Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., 2017. *Enriching word vectors with subword information*. <https://arxiv.org/abs/1607.04606v2>
- Cesconi, F., 2017. *What is the main difference between word2vec and fasttext?* <https://medium.com/@federicocesconi/what-is-the-main-difference-between-word2vec-and-fasttext-57bdaf3a69ef>
- Chorowski, J., Bahdanau, D., Serdyuk, D., Cho, K., Bengio, Y., 2015. Attention-based models for speech recognition. In *Proceedings of the 28th International Conference on Neural Information Processing Systems* - Volume 1, NIPS'15, pages 577–585. 2, 24.
- Davydova, O., 2017. *7 types of artificial neural networks for natural language processing*. <https://medium.com/@datamonsters/artificial-neural-networks-for-natural-language-processing-part-1-64ca9ebfa3b2>
- Du, C., Huang, L., 2018. Text classification research with attention-based recurrent neural networks. *International Journal of Computers Communications Control*, 13:50, 02. xiii, 25, 26, 32
- Elsheikh, A., Yacout, S., Ouali, M-S, 2018. Bidirectional handshaking lstm for remaining useful life prediction. *Neurocomputing*, 323, 10. 21
- Facebook Inc., 2020. *Word vectors for 157 languages*. <https://fasttext.cc/docs/en/crawl-vectors.html>.
- Firth, J. R., 1957. [https://en.wikipedia.org/wiki/John\\_Rupert\\_Firth](https://en.wikipedia.org/wiki/John_Rupert_Firth).
- Fombrun, C., van Riel, C., 1997. The reputational landscape. *Corporate Reputation Review*, 1:5-13.
- Galassi, A., Lippi, M., Torroni, P., 2019. Attention, please! A critical review of neural attention models in natural language processing. *arXiv:1902.02181 [cs.CL]*
- Gatzert, N., 2015. The impact of corporate reputation and reputation damaging events on financial performance: Empirical evidence from the literature. *SSRN Electronic Journal*, 01.
- Jiang, Q., Chen, L., Xu, R., Ao, X., Yang, M., 2019. A Challenge Dataset and Effective Models for Aspect-Based Sentiment Analysis. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 6280–6285, Hong Kong, China, November 3–7. Association for Computational Linguistics, DOI: 10.18653/v1/D19-1654.
- Kachaniuk, O., 2019. Corporate Reputation Estimation based on biLSTM Neural Network Sentiment Analysis with Attention Mechanism. *Master thesis, VU Vrije Universiteit Amsterdam*.
- Keh, H.T., Xie, Y., 2009. Corporate reputation and customer behavioral intentions: The roles of trust, identification and commitment. *Industrial Marketing Management*, 38:732-742.
- Koehrsen, W., 2018. *Neural network Embeddings Explained*. <https://towardsdatascience.com/neural-network-embeddings-explained-4d028e6f0526>.
- Lai, S., Xu, L., Liu, K., Zhao, J., 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*. 20
- Lester, A., 2009. Growth Management: Two Hats are Better than One. *Springer*.
- Lurz, K.-K., 2018. Natural language processing in artificial neural networks: Sentence analysis in medical papers. *Student paper, Lund University*. <https://lup.lub.lu.se/student-papers/search/publication/8948103>.
- Manning, C., Socher, R., 2017. *Word vector representations: word2vec*. <https://www.youtube.com/watch?v=ERibwqs9p38>.
- Nicholson, C., n.d. *A Beginner's Guide to Attention Mechanisms and Memory Networks*. <https://pathmind.com/wiki/attention-mechanism-memory-network>.
- Pennington, J., Socher, R., Manning, C., 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.
- Perone, C.S., Silveira, R., Paula, T.S., 2018. Evaluation of sentence embeddings in downstream and linguistic probing tasks. *arXiv:1806.06259 [cs.CL]*.
- Qian, P., Qiu, X., Huang, X., 2016. Investigating language universal and specific properties in word embeddings. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, August. 22
- Rajasekharan, A., 2017. *What is the difference between fasttext and glove?* <https://www.quora.com/What-is-the-difference-between-fastText-and-GloVe>
- Ross, J.K. III, Patterson, L.T., Stutts, M., 1992. Consumer perceptions of organizations that use cause-related marketing. *Journal of the Academy of Marketing Science*, 20:93-97, 12.
- Silvija, V., Ksenija, D., Igor, K., 2017. The impact of reputation on corporate financial performance: Median regression approach. *Business Systems Research*, 8(2):40-58.
- Speaks, J., 2018. Theories of meaning. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter edition. 22
- Vocabulary.com, n.d. *Meaning*. <https://www.vocabulary.com/dictionary/meaning>.
- Wang, F., Jiang, M., Qian, C., Yang, S., Chenxi Li, C., Zhang, H., Wang, X., Tang, X., 2017. Residual attention network for image classification. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6450–6458. 24
- Wikipedia, n.d. *Word*. <https://en.wikipedia.org/wiki/Word>.